(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/88088 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/14827

(22) International Filing Date: 16 May 2001 (16.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/577,408 18 May 2000 (18.05.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/577,408 (CIP)
Filed on 18 May 2000 (18.05.2000)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way, #14, San Jose, CA 95117 (US). **DRMANAC, Radoje, T.** [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/88088 A2

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

626	8677	A	1582	2	1296	ALCEPQPFQGSGCVIAILGRKMFSSVAHL ARANPFNTPHLQLVHDGLGDLRSSSPGP TGKPRRPSQ/HMAAAPVEEQYSCDYGSG RFFILCGLGGIISCGTTHTALVPLDLVK\C RMQVDPQKYKG\IFNGFSVTLKEDGV\R GLAKGWAPTFLGYSMQGLCKFGFYEVF KSL\YSNMLGE\ENTYL*RTSLYLAASAS\ AEFFADIALAPMEAAKVRIQTQP\GYANT *EGISFPKCIKEEGLTSILQGGLLPLWMR QIPYTMN*SSPCLERTV\EA\LYKFV\VPK\ PRRE*FKRQSRLVVTIW*QVTIARVFCAN CFSPLPEFLG*PVLD*GKKVSQCFLWVLQ RDLGFK\GV\WKGLFA\RII\MIGT\LT\ALQ WFIYYSVKGYFR\LPRP\PPP\EMQES\LKK KLGVNSVVRIKANCGLNLLVDPVFEESA KGTFIYLTV
627	8678	A	1583	127	433	RPLESWIGLVRCNICRSPIAEAVFRKLVT DQNISKNWRVDSAATSGYEIGNPPDYRG QSCMKRHGIPMSHVARQ\DLNRKSNRV KTCKAKIELLGSYDPQKQL
628	8679	A	1584	2	535	
629	8680	A	1585	551	1299	PADPPRPSYYRHRTPPQAHWSRLRRSRL RRRGSHTRCPVGVGAGLRRRAGARLAV RLRASACGTPRCLGASARGKMAEQATK SVLFVCLGNICRSPIAEAVFRKLVTDQNI SKN/WEGRQRGNFRWVIDSGAVSDWNV GRSPDPRAV\SCLRNHGIHTAHKARQIT\ KEVFPTFDYILCMDESNL\RDLNRKSN\R VKTCKS*KFELPWEL*SPQKQLIIED\PYY GE*LWTLETVYQQ\CVR\CCRAFL\EKAH
630	8681	A	1586	1	1239	
631	8682	A	1587	298	408	
632	8683	C	1588	92	244	MRCEIILVLIPYVYFYSNKLLCSRLXXXX XGGAVLKNPWGGQSLPGLAR**
633	8684	A	1589	33	191	RDDPRVRPPPNSHT*PQQEPGL*LIKCTSP PQAPAPRTVHGPYFYMRLIKMF
634	8685	A	159	445	673	RECLH*PRMATQRKHLVIDFNAYITCYIC KGYLIKPTTVT\ECLHT/FCRCMEAFPSLL LA
635	8686	A	1590	3	1285	
636	8687	A	1591	3	3469	QPGHTIYLLPTVVICNLLPCELDFYVKGM PINGTLKPGKEAALHTADTSQNIELGVSL ENFPLCKELLIPPGTQNYMVRMRLYDVN RRQLNLTIRIVCRAEGSLKIFISAPYWLIN KTGLPLIFRQDNAKTDAAGQFEEHELAR SLSPLLFCYADKEQPNLCTMRIGRGIHPE GMPGWCQGFSLDGGSGVRALKVIQQGN RPGLIYNIGIDVKKGRGRYIDTCMVIFAP RYLLDNKSSHKLAFAQREFARGQGTA
637	8688	C	1592	398	655	MMFPLAFSLPLKNAFHISVCRVCPGYTG FAKRALTALNLDTSLSANCCNTPPAEXP NVHNPCYMGLSKPARXSKLGSMCKGSS XH*
638	8689	A	1593	1	930	
639	8690	A	1594	1	134	

640	8691	A	1595	3	2455	HASVCPAVGVQRLCLFPCVSLQALFMGS PLRFDGRFF\LVTGAGAGLGRAYALAFA ERGALVVVNDLGGDFKGVGKGSLAADK VVEEIRRRGGKAVANYDSVEEGDKVVK TALDAFGRIDVVVNNAGILR/DINSFARIS DEDWDIIHRVH\LRGSFQVTPAAWEHMK KQKYGRSIMTSSASGIYGNFGQANYSAA KLGLLGLANSLAIEGRKSNIHWNTIAPNA GSRMTQTVMPEDLVEALKPKYVAPLVL WLCHQSCEENGGLFEVGAGRIGKLRWE RTLGAIVRQKNHPMTPEAVKANWKKIC DFENASKPQSIQESTGSIIEVLSKTDSEGG VSANYTSRATSTATSGFAGAIGQKLPPFS YAYTELEAIMYALGVGASIKDPKDLKFI YEGSSDFSCLPTFGVIIGQKSMMGGGLA EIPGLSINFAKVLHGEQYLELYKPLPRAG KLKCEAVVADVLDKGSGVVIIMDVYSY SEKELICHNQFSLFLVGSGGFGGKRTSDK VKVAVAIPNRPPDAVLTDTTSLNQAALY RLSGDWNPLHIDPNFASLAGFDK\PILHG\ LCTFGIFCQGVLLQQFCR*MDVVQGFKG N*RARF\AKPVYPGANFYQT*ECWKE\G NRNSFFKPKVQGNLETLVISKWHMWDL GTQHSGYFSLRTPSEGPGSFRVPLVFEE\I GRRLKDIG\PEVVK\KVNAVF\EWHITKG GNI\GAKWTIDLK\SGSWEKLYQGPS/KK GAADTTIH/ILSDEDF/LWEVVLGQA*PSR KAFFSGPG*RPQGGTSMA*AQ\KLSDGFL KDYAKLLKGTPTLLIKMESIKIPPPHPQIC LDYSAKS
641	8692	A	1596	2	289	
642	8693	A	1597	1	397	
643	8694	A	1598	1	410	STMISPVLILFSSFLCHVAIAGRTCPKPDD LPFSTVVPLKTFYEPGEEITYSCKPGYVS RGGM\RKFICPLTGLWPINTLKCTPRVCP FAGNLRKMGAVRLITDFLNYSPTRFSFSL LTWGFILEWALDS\AKCIEGG
644	8695	A	1599	19	1215	CQCDSSTMIFSRCSSLFSSFLCHVAIAGRT CPKPDDLPFSTVVPLKTFYEPG\EEITYSC KPGYVSRGGIEESLSCPL\TGTVGPFNTSG NVTPRVCPF\AGIFRKMGGRTLITTF*NYP NTDPVFSLLTLGF*FWNGALDFWPSCTG GKGKW\SP\ELPGLVAPII\CPP\PSIP/TGFA TLHVLLRPFRLGNNSPPIGDTAVFECLAH NMAMFG\NDTIT\CTTHGKLDLNYPECR GSKMPPFPHQDPDNGIW*TYPCQNPNTL FTRVKAPHLGLPHDGIFSGMGPRKE\EC *PQTWGKPGSWPLAPSW*KPSLVKGTPV KKRPTVV/YPQGERVKDSREKFKEWECL HG**KFLSFCKNKEKKCSYTEDAQCIDG TIEVPKCFK\EHSSLAFWKT\DAS\DVKPC
645	8696	A	16	3	145	SSSSSDFAGQTL*STQTVQN*FKKVLKPG RLYPVPIATMGIKEPLIS
646	8697	A	160	22	849	WIERDLLNCIKRLK/PTTNNMLNDEIVNIS PKIIKIRQGYLLSMILFGIVQKDLTRKLM QGRETKGIEIRKEVKL*KRKRI*ISICRCH E*IW*VPCIKVMQKAFYDIPAKNMENEIL KKQCHFKDPSSA*REKMRLICFEELYPEN KITKEERDRI/RRTISKLLLFPKFHLQP*NP RQVSLMLN*QANF*EFICIFQKS\KIVKAI L*NGQRGLKFLNIKTCYKAIEIMKVLIWH KD\KKLD*WNSIQVSKVDPRVYHHLSFE KGDIEV*WGKGCSFQ
647	8698	A	1600	1	282	

648	8699	A	1601	1	453	EFGSQQLGRREEWQRQGSPVSRRLSARR GPQAPGTRLPRRHPARAFPAATMPKRKV SSAEGAA*LEPNSRSARLSAKPPAKGEA KPKKAAAKDKSSDKK\VQTKGKRGAKG KQ\AEVANQETKEDLPAENGETKTEESP\ ASDEAGEKEAKSD
649	8700	A	1602	146	824	TWGKGDPKKPRGKMSSYAFFVQTCR\EE HKKKHPDASVNFS/ESFSKKCSERWKTM SA*R/EKGKFEDMAKA\DKARY\EREMK TYIPPQRGRQKRKFKDSQLHPRGPPSGLL SSSCSEYRPKIK\GEHP\GL\SIGDVAKKLG RDVGINTAAD\DKQPYEKK\AAKLKEKY EKDIAAYRAKGKPDAAKKG\VVKAEKS KKKKEEEEDEEEG\DEEDEEEEEDEEDEE DEEEDER
650	8701	A	1603	1	223	
651	8702	A	1604	1	400	FADD/PSDK/FFTSNNGMQFSTGHNDND KFEGNCAEQDGSGWWMNKCHAGHLNG VYYQGGTYSKASTPNGYDNGIIWATWK TRWYSMKKTTMKIIPFNRLTIGEGQQHH LGGAKQVRPEHPAETEYDSLYPEDDL
652	8703	A	1605	18	365	NILIKVYFNSKNDFKIFHELFFKQNYMKN MYKSVINVIDIFMNKFQ/SEKYPII/DKGS LNK*MLTILALKSNTTVRLIRDTAFYYVR EHIINVSSKRARYWVCVGFI*ASC*QPPL F
653	8704	A	1606	212	1645	HYKARSSGHSDIMSWSLH\ARNLILYFY ALLFLSSTCVAYVATRDNCC\ILYERFGC YC\PTTCGIADFLSTYQTRVD*DLQSL\ED ILHQVENKTS\EVKQLIKAIQL\TYNPD\ES SKPNMIDAATLKSRKMLEEIMKYEAS\IL THDSSIRYLQEII*FQIIQKIVNL\KEK\VAQ LEAQCQEPCKDTVQIHDITGKDCQD\AN KGAKQSGLYFIKPLKANQQFLVY\CEIDG SG\NGWTVFQKRLDGSVDFKK\NWIPYK EGFGHLSPTGTTEFLAGEMRKIHFD*GTQ S\AIPYGI*GVGTGKTWEWARNQYCRSM PLFKVVHEVD\KYRFTYAYFAGGDAEDA FDGYDFG\DDPSDKFFHIPIMAMQFTYLG TMDNDKV*KANCA*/QQGWDPGWWDG NKC\HAG\HSSMGVLFTQGWALYFQKAS YLPNGLWIMGIIWA\TWK\TRWVFR*RKP TMKIIP\FNRLTIGEGQQH\HLGGSQTGLE TF
654	8705	A	1607	2	529	GTVAACGACYWLLGLMAVRASFENNCE IGCFAKLTNTYCLVAIGGSENFYSVFEGE LSDTIPVVHASIAGCRIIGRMCVG\TEEIL ADVLKVEVFRQTVADQVLVGSYCVFSN QGGLVHPKTSIEDQDELSSLLQVPLVAG TVNRGSEVIAAGMVVNDWCAFCGLDTT STELSVVE
655	8706	A	1608	18	889	GVQGTVAACGACYWLLGLMAVRASFE NN\CEIGCFAKLTNTYCLVAIGGSENFYS VFEGELSDTIPVVHASI\AGCRI\IGRMCV GN\RHGLL\VPNNTTDQ\EL\QHISATGLP RHSGRFRAGWKERFLSLWGNFFNHLAID YVGLGSNQD\LDKGRQEEISGQMLFKGW EVFRQTV\ADQV\LVES\YCVFSNPGRAW VPSPRPFQ*RPRNELSSISFKVPL\VAGTC* TKGSEVICLLGMGGEMNWCA\FCGPGTP NPAQSCQVVEECLQS*NEAPALAPIANR ACGNSL\IDSLT
656	8707	A	1609	1	248	GPLIWEWPASPEPPPLPWGKPRMQ/SG*Y G*TP*IPKIRFPKKPFPPFPQALEPQQKGP N*AHP*EPTPAKKYSPQRVQKVPK

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-8051, a mature protein coding portion of SEQ ID NO: 1-8051, an active domain of SEQ ID NO: 1-8051, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

(a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-8051.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-8051, a mature protein coding portion of SEQ ID NO: 1-8051, an active domain of SEQ ID NO: 1-8051, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO: 1-8051, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 8052-16102, the mature protein portion thereof, or the active domain thereof.

21 The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprises the sequence information of at least one of SEQ ID NO: 1-8051.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

28 A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.